

A Meta-analysis of Within-Household Selection Impact on Survey Outcome Rates, Demographic Representation and Sample Quality in the European Social Survey

Piotr Jabkowski

Adam Mickiewicz University in Poznan

Face-to-face surveys of the general population often start with a probability sample of the households or addresses. However, even if a probability sample of households or addresses has been drawn, random selection of the target respondent within the selected household is crucial for obtaining a probability sample of individuals comprising the population. Over a dozen procedures of within-household selection have been described in survey literature. This article is concerned only with the two most popular of these, i.e., with the Kish grid procedure and the class of birthday procedures (i.e., next-birthday, last-birthday and closest-birthday methods). The main goal of this paper is to address the question of whether the Kish grid and birthday methods differ in their impact on: (1) refusal and cooperation rates, and (2) demographic representation of the survey sample, as well as (3) the degree of interviewers' influence on the selection process. Based on 98 different surveys from all seven rounds of the European Social Survey, a meta-analysis was conducted to generate quantitative measures indicating the size of the overall impact of Kish grid and birthday procedures. Several conclusions can be formulated based on the analysis. Firstly, Kish grid samples (compared to birthday samples) result in significantly higher odds of receiving refusals and significantly lower odds of obtaining cooperation. Secondly, both Kish grid and birthday samples have a similar and significant impact on gender and age imbalance. Finally, birthday procedures give interviewers greater opportunity to influence the selection process. The latter means that the use of Kish grid samples is usually associated with a higher quality of the within-household selection, even though in such samples refusal rates are significantly higher and cooperation rates are significantly lower compared to those obtained in birthday samples.

Keywords: probability samples; within-household selection; Kish grid; birthday procedures; internal and external criteria of representativeness; European Social Survey.



Acknowledgements

The author thanks two anonymous reviewers for their comments that substantially improved the paper.

1. INTRODUCTION

Face-to-face surveys of the general population often start with a probability sample of households or addresses (Stoop, Billiet, Koch & Fitzgerald, 2010). This is a consequence of the available sampling frames and has a significant impact on sample quality and its fieldwork realization (Menold, 2014). While the frame of individuals allows a direct random sampling of population units, the implementation of address and household samples renders necessary indirect selection of individuals comprising the population. However, even if a probability sample of addresses and households has been drawn, it does not automatically lead to a probability sample of individuals. In fact, the random selection of target respondents within selected households is crucial for obtaining a probability sample of individuals, and for making inferences from address and household samples to the target population.

Survey literature describes over a dozen procedures for within-household selection of a target person (Gaziano, 2005). Some of them are probabilistic, while others are quasi-probabilistic or even non-probabilistic. Kish grid (Kish, 1959), age-order and age-only procedures (Denk & Hall, 2000), as well as the so-called full-enumeration technique (Denk & Hall, 2000) are examples of probabilistic procedures of within-household selection. In turn, next-birthday, last-birthday or closest-birthday methods (Salmon & Nichols 1983) are examples of quasi-probabilistic procedures. Finally, the Trolldahl-Carter technique (Trolldahl & Carter, 1964) with its various variants, such as the Paisley and Parker Stanford modification (Paisley & Parker, 1965), Bryant's correction for too many females (Bryant, 1975), Groves and Kahn's modification (Groves & Kahn, 1979), the Czaja-Blair-Sebestik procedure (Czaja, Blair & Sebestik, 1982), the Hagan-Collier "alternative" (Hagan & Collier, 1983), the youngest male / oldest female (YMOF) method (Keeter & Fisher, 1997) and the most recent technique called the Le-Brick-Diop-Alemadi method (Le, Brick, Diop & Alemadi, 2013) are the best-known examples of non-probability procedures.

Within-household selection procedures differ not only in terms of their basic assumptions, but also in terms of fieldwork implementation and the survey errors resulting from each selection method. For example, probabilistic procedures require a list of all persons living in the household. Quasi-probabilistic procedures bypass the necessity of listing household members, however this is at the cost

of incomplete randomness of the selection process. Finally, non-probabilistic procedures rely on quota sampling. Their basic purpose is to achieve demographic representation of gender and age, even if they do not preserve the distributions of other variables. This is the primary reason why non-probability procedures of within-household selection are not recommended and are avoided whenever possible in surveys based on probability samples.

This article is concerned with two methods of within-household selection only, i.e., the Kish grid and birthday procedures (in next-birthday, last-birthday and closest-birthday variants). A meta-analysis of ESS data was used to compute an 'overall effect size' (Borenstein, Hedges, Higgins & Rothstein, 2009), i.e., to provide quantitative measures of the impact of the Kish grid and birthday methods on demographic representation and selection quality. However, the first goal of this paper is to verify whether birthday methods, believed to be less time consuming and less intrusive for household members, result in lower refusal rates and higher cooperation rates. The second goal is to verify whether the Kish grid and birthday procedures differ in their impact on demographic representation of gender and age. Both gender and age imbalance was evaluated by means of external criteria of representativeness, i.e., by comparing survey results with more reliable sources of external benchmark data (Koch, 2016). The third goal of this paper is to address the question of whether the Kish grid and birthday procedures differ in their impact on the degree of interviewers' influence on the within-household selection process. This is a very serious problem, since the limited scope for effective monitoring of the quality of fieldwork implementation of within-household selection procedures increases the risk of illegitimate substitution of another unit for a nonresponding one (not present at home and/or reluctant to cooperate). It should be noted that the presence of significant interviewers' impact on the selection process was verified in this article by means of the so-called internal criteria of representativeness (Kohler, 2007; Sodeur, 1997) with a form of inference proposed by Menold (2014).

The first section of this article provides the theoretical and empirical background of the Kish grid and birthday methods. Hypotheses are described in the second section, while the third section provides information about: (a) the method, (b) studies included in meta-analysis, (c) the idea of external and internal benchmark data and (d) analytic procedures used for estimating the impact of the within-household selection method on demographic representation and selection quality. Results are presented in the fourth section, while the last section of the article contains conclusions.

2. THEORETICAL AND EMPIRICAL BACKGROUND OF THE KISH GRID AND BIRTHDAY METHODS OF WITHIN-HOUSEHOLD SELECTION

The Kish grid (Kish, 1949) constitutes one of the best-known examples of a probabilistic procedure for within-household selection. It is carried out by the interviewer who is responsible for listing all household members by sex, and within each sex grouping by age (from oldest to youngest). After enumerating all eligible units, the interviewer uses two sequences of numbers. The first sequence of numbers corresponds to the possible size of the household (starting from 2, up to usually 12 members), while the second sequence is unique for each household and indicates the number assigned to a target respondent according to the size of household. The second sequence contains randomly selected numbers, so that the selection process may be purely random. Of course, if there is only one eligible person in a household, he or she is the target person and no within-household selection is made.

The main advantage of the Kish grid lies in its capacity to produce consistent and unbiased estimates (Yan, 2009); that there is a lower opportunity for interviewers to interfere with the process of selecting or excluding some individuals is also pointed out as an advantage (Smith et al., 1995). One should note that the Kish grid requires compilation of a full inventory of all eligible individuals comprising a household, which is time consuming and may be extremely intrusive for household members. Thus, as suggested by Gaziano, “although probability methods are preferable, they often increase nonresponse. Quasi-probability and nonprobability techniques have been devised to increase cooperation and decrease costs, although they sacrifice the advantages of randomness” (Gaziano, 2005, p. 124).

Quasi-random procedures avoid a full inventory of household members. Among the most popular methods of quasi-random selection one finds those using birthdays, i.e. the last-birthday, next-birthday or closest-birthday methods. The process of within-household selection with the last-birthday method involves asking to make an interview with the eligible household member who had the most recent birthday. Similarly, the next-birthday method consists of interviewing the household member who will have the next birthday, while the closest-birthday method is based on interviewing the household member whose birthday is closest to the date of interview, regardless of whether this date is in the past or in the future.

Although early studies of birthday procedures demonstrated the randomness of population birthdays (Salmon & Nichols, 1983; O’Rourke & Blair, 1983), the results of subsequent analyses performed by Groves & Lyberg (1988) put the randomness of population birthdates into serious doubt. An additional essential insight into the problem of birthday procedures was provided by Lavrakas, Bauman, & Merkle

(1993) and Lavrakas, Stasny, & Harpuder (2000). The results of both these studies demonstrated that birthday procedures led to higher cooperation rates and lower costs (in comparison to the Kish grid); however, the procedure also entails multiple dangers stemming both from measurement errors and interviewer impact on the selection process. Both teams led by Lavrakas (1993, 2000) demonstrated that one of the main faults of birthday procedures consisted in sample-selection of the very person with whom the contact was made, irrespective of whether their birthday was in fact last, next or closest to the date of successful contact; such incorrect selection took place in approximately 20 percent and more of all cases. In turn, the experimental studies conducted by Smith et al. (1995) demonstrated that the Kish grid has been successfully implemented in 93% of all selected households, while in 99% of all cases the interviewers adequately used the rules of within-household selection. In other words, the advantages of birthday procedures over the Kish grid, i.e., higher cooperation and lower refusal, were nullified by a distortion of the selection process.

It should be pointed out, however, that the analyses made by Lavrakas et al. (1993, 2000) and Smith et al. (1995) were based on telephone surveys with random digit dialing. In consequence, their findings may be mode-specific, since the challenges one has to face in the process of within-household selection of a target respondent are somewhat different in telephone-administered and face-to-face surveys. For example, in telephone surveys the Kish grid enumeration of all eligible population units within the household is much more troublesome; it is harder to gain detailed information about the household members by telephone and it is much easier break-off contact on the telephone than in the case of face-to-face interaction. Since the analyses presented in this paper are based on the ESS project, where only face-to-face interviewing is allowed by protocol, they offer the possibility of verifying whether the findings presented on the basis of telephone surveys hold true for face-to-face surveys.

One should also bear in mind that the within-household selection process poses another problem, i.e., that of who should be classified as a household member. In fact, the process of within-household selection may lead to significant under-coverage of eligible population units and it may result in the so-called within-unit coverage bias (Ziniel, 2008). As concluded by Groves, "In addition to the problems of attempting to fix the population in time, however, new sources of coverage problems that result from the listing process arise. Listing of persons is not based solely on the interviewer's observational activities but usually requires the questioning of a household member. The level of coverage within households is thus dependent on the behaviour of the interviewer and the household member supplying the information" (Groves, 1989, p. 109-110). The impact of within-household selection procedures on under-coverage rate and coverage bias has

been analysed in many methodological studies. For example, Martin (1999) and Gaziano (2008) showed that the informants tend not to mention residents who stay in a household most of the time, but not all of the time. Another problem is under-coverage of certain kinds of household members, especially young males, older females, poorer or less educated people (Brogan et al., 2001; Brooks & Bailar, 1978; Korns, 1977; O'Rourke & Lakner, 1989; Valentine, Betty, & Valentine, 1971).

One may also note that even the way of recording information about household members may have an impact on the under-coverage rate. For example, Tourangeau, Shapiro, Kearney and Lawrence (1997) demonstrated that differences in formulation of the question pertaining to enumeration of the household members affect the coverage of population units. The authors compared three versions of roster questions. The first version was the control and began by asking informants to name all the persons living in a household, as in a standard version of the Kish grid. Two other versions of roster questions were experimental and began by asking how many persons had spent the previous night at the household; the first version used full names, while the second used initials instead of full names. The results indicated that the two experimental versions identified more persons per household than the standard version of Kish grid questions. However, only the version with initials yielded more persons identified as permanent residents of household.

Even a brief overview of survey literature demonstrates that each of the within-household selection methods has its advantages and disadvantages. Some of the procedures are easy to implement, are low-cost, are not time consuming and are non-intrusive, while others require much more effort on the interviewers' part, more time and more costs. The price of the first is usually the loss of full randomness of the selection process, while the cost of the latter is an increase in the number of refusals to participate in the survey (Lavrakas et al., 1993). Thus, the final decision about which within-household selection method to use in a sampling process should always arise from the total survey error paradigm. The choice of the best procedure of within-household selection constitutes a major challenge especially in cross-country comparative surveys, as one essentially expects their sampling procedures to imply similar random and non-random errors (Lynn, Gabler, Häder, & Laaksonen, 2007).

3. RESEARCH HYPOTHESES

As was shown in the introduction, the following three main problems are subject to empirical analysis in this article. Firstly, the impact of Kish grid and birthday procedures on refusal rates and cooperation rates are considered. Secondly, the impact of Kish grid and birthday procedures on gender and age imbalance are

analysed. Finally, the impact of Kish grid and birthday procedures on the quality of the within-household selection process, i.e., presence or absence of interviewers' influence on the selection process, is investigated. Hypotheses describing the possible impact of Kish grid and birthday procedures on survey outcome rates (H1), demographic representation (H2), and interviewers' influence on the selection process (H3) are presented below.

3.1. Hypothesis regarding the impact of Kish grid and birthday procedures on survey outcome rates

As was mentioned in the previous sections of this paper, the implementation of the Kish grid requires a full enumeration of all eligible population units living in a household. The belief that such enumeration has an intrusive character is common in survey literature (O'Rourke & Blair, 1983; Yan, 2009); only Gaziano (2005) has pointed out that the Kish grid may not be as intrusive as was previously believed by researchers. However, even Gaziano (2005) concluded that birthday methods have the advantage of higher cooperation rates over the Kish grid. Most of the critics of Kish grid sampling emphasize a significant lowering of cooperation rates through an increase in refusal rates as a direct consequence of such intrusive enumeration of all household members (Binson, Jesse, & Catiana, 2000; Oldendick, Bishop, Sorenson, & Tuchfarber, 1988). Yan (2009) observed, in turn, that the odds of obtaining complete interviews were smaller for the Kish grid than for the birthday methods. Moreover, the Kish method results in a significantly higher chance of eliciting refusals from respondents than in the case of the less intrusive last-birthday method. In the analysis of O'Rourke & Blair (1983) one can also find the conclusion that the higher refusal rates obtained by the Kish grid are caused by refusals occurring at the phase of selection of respondents, which confirms the intrusive character of listing the household members.

The hypothesis describing the impact of the Kish grid and birthday procedures on survey outcome rates is therefore as follows:

Hypothesis 1 (H1): Birthday procedures of within-household selection, as they are less intrusive than Kish grid enumeration of all household members, result in significantly lower refusal rates and significantly higher cooperation rates.

3.2. Hypothesis regarding the impact of Kish grid and birthday procedures on the demographic representation of a survey sample

The results of many empirical studies indicate that both the Kish grid and birthday procedures produce an overrepresentation of women (Denk, Guterbock, & Gold, 1996; Gaziano, 2005; Yan, 2009); only a few found no evidence of a significant

impact of Kish grid and birthday methods on gender imbalance (O'Rourke & Blair, 1983). Moreover, most empirical analyses indicate that Kish grid sampling leads to a smaller overrepresentation of women. For example, Yan (2009) revealed that the odds of selecting females with the Kish grid are smaller than the odds of selecting females with last-birthday methods. Similar conclusions were also formulated by Denk and Hall (2000), whose analysis indicated that the Kish grid worked best towards reducing the overall inclusion of women in two- and three-adult households. One may also expect a significant impact of Kish grid and birthday procedures on age representation. Both the Kish grid and birthday procedures tend to underrepresent young and old people, however, samples based on the Kish grid contain a greater number of older, and fewer younger respondents (O'Rourke & Blair, 1983).

The hypothesis describing the impact of the Kish grid and birthday methods on gender and age imbalance is therefore as follows:

Hypothesis 2 (H2): Both the Kish grid and birthday methods have a significant impact on gender and age imbalance, i.e. they result in an overrepresentation of females, and an underrepresentation of younger and older respondents. However, this impact is less significant in Kish grid samples.

Of course, some part of the gender and age imbalance observed in Kish grid and birthday samples may be caused both by the selection method and by the differentiation between gender and age groups in the propensity to respond. As the nonresponse impact on gender and age imbalance has not been controlled in the studies referred to above, it may be assumed that not only the selection method, but also the nonresponse process may have had an impact on the deviations observed in these studies. It should be noted, however, that the analysis presented in this paper also suffers from being conditional on survey participation; the nonresponse effect is partially controlled only when the following hypothesis (H3) is verified.

3.3. Hypothesis regarding the impact of Kish grid and birthday procedures on the degree of interviewers' influence on the within-household selection process

Birthday methods appear to involve more problems with the accuracy of within-household selection than the methods requesting the listing of respondents by name (Gaziano, 2005). The interviewers' degree of freedom in selecting an adult different from the one with the last, next or closest birthday is much higher than with Kish grid, where age order enumeration of all household members must be completed before the selection starts, and the identification number of the target person within a given household is predetermined. The risk of illegitimate

substitution is considerable especially when a person with the last, next or closest birthday is absent from home or reluctant to participate in the survey. In turn, a full enumeration of all the household members in Kish grid samples reduces the interviewers' freedom to choose a person different from the one that should be selected. This means, however, that interviewers' influence on the selection process may pose a greater problem in samples utilizing birthdays.

Therefore, the hypothesis describing the impact of Kish grid and birthday methods on the degree of interviewers' influence on the within-household selection process is as follows:

Hypothesis 3 (H3): In birthday methods the degree of interviewers' influence on the selection process is much higher than in Kish grid samples.

4. METHODS

4.1. Database

The hypotheses presented above will be verified on the basis of results from all seven rounds of the European Social Survey. The ESS project is a well-known cross-country survey that has been conducted biennially since 2002. The target population contains "all persons 15 years or older resident in private households within the borders of the nation, regardless of nationality, citizenship, language or legal status" (Lynn et al., 2007, p. 109). Depending on the available sampling frames, one of three types of sample are utilized in the ESS project, i.e., (1) individual name samples, (2) household samples or (3) address samples (Stoop et al., 2010). Both in the address and household samples the final selection of a target respondent is always performed by a strictly random (Kish grid) or quasi-random (last- / next- / closest-birthday) procedure of within-household selection. The number of participating countries varies from 20 to more than 30 countries in each ESS round. However, in this article not all the surveys were taken into consideration, but only those which implemented Kish grid or birthday methods in the sampling process (Table 1). Altogether, 98 different surveys from 25 countries were included in the meta-analysis. In 34 cases the Kish grid was utilised, while 41 cases used the last-birthday method, 18 the next-birthday and only 5 cases used the closest-birthday method.

Table 1. Classification of ESS countries^a with respect to within-household selection methods

ESS round	Kish grid	Birthday methods
ESS1_2002	CH, CZ, GB, GR, IT	Last-birthday: ES, FR, IL, LU, PT Next-birthday: AT, IE, NL Closest-birthday: -
ESS2_2004	CH, CZ, GB, GR, IT, TR	Last-birthday: FR, LU, NL, PT, UA Next-birthday: AT, IE Closest-birthday: -
ESS3_2006	CH, GB, NL, RU	Last-birthday: BG, CY, FR, PT, UA Next-birthday: AT, IE, LV ^b Closest-birthday: RO
ESS4_2008	CH, GB, GR, NL, RU, SK, TR	Last-birthday: BG, CY, FR, IL, LT ^b , PT, UA Next-birthday: AT, HR, IE, LV Closest-birthday: CZ, RO ^b
ESS5_2010	BG, GB, GR, NL, RU, SK	Last-birthday: CY, FR, IL, LT, PT, UA Next-birthday: AT, HR, IE Closest-birthday: CZ
ESS6_2012	FR, GB, NL, RU, SK	Last-birthday: AL, BG, CY, IE, IL, LT, PT, UA Next-birthday: CZ Closest-birthday: XK
ESS7_2014	GB	Last-birthday: IE, IL, LT, NL, PT Next-birthday: CZ, FR Closest-birthday: -

^a Countries are labelled according to ISO31166-1

^b Data from Latvia (ESS3), Lithuania (ESS4) and Romania (ESS4) were excluded from meta-analysis since design weights have not been produced.

One should also bear in mind that ESS puts a strong emphasis on the standardisation of the sampling process and fieldwork procedure in such a way as to enable cross-country comparisons of results despite utilisation of different types of samples. For example, each of the ESS National Coordination teams is supported by a member of the ESS Sampling Expert Panel in the process of choosing a sample design suitable for implementation in each country. The sampling process must be finally approved by the whole ESS Sampling Expert Panel before fieldwork starts in order to ensure that it is comparable with those utilised in other countries (Stoop et al., 2010). This is a very desirable feature in meta-analysis of the ESS data because one can expect that the impact of Kish grid and birthday procedures on survey outcome rates, demographic representation and survey quality is determined by the different character of within-household selection methods. This is due to the fact that the sampling process and fieldwork execution of survey samples are strictly standardized in each country participating in the ESS project.

Nevertheless, there are also some serious limitations to the analysis presented in this paper that arise from some specifics of the ESS project. Firstly, within-household selection procedures were not randomly assigned to countries. This, in fact, may have increased the likelihood of reporting a partly spurious effect of within-household selection procedures, and makes it hard or even impossible to extend any general conclusions resulting from the analysis presented in these studies to other studies. In other words, the findings are indicative of a possible effect of within-household selection method on survey outcome rates and selection quality in the ESS project, but they are far from general or conclusive. Secondly, within this analysis each of the country-level samples is assumed to represent a certain type of within-household selection procedure, which means that the assumption is made that it is this type of procedure that constitutes the primary source of differentiation within the survey outcomes, demographic representation and selection quality. However, many other factors may contribute to such differentiation. From a practical point of view, the most interesting factors are those which remain under the control of the researcher. Apart from the types of within-household selection procedures in address and household samples, these are typically considered to include type of sampling frame, mode of data collection, rules for conversion of refusals, the presence of incentives, and interviewer training and supervision (de Heer, 1999). Nonetheless, the observed differentiation can also result from factors remaining outside the researcher's control, such as survey climate (Bethlehem, Cobben, & Schouten, 2011; Groves & Couper, 1998; Smith, 2007). Thirdly, while Kish grid sampling is purely random, birthday sampling is only quasi-random. Even though the Kish grid and birthday procedures offer opportunities for performing the selection of a target respondent in a way that is completely independent of the researcher, the interviewers, the household members and the selected individual, from a statistical point of view it is not so obvious that the sample of population units drawn under quasi-random procedures permit an application of the methods of statistical inference. To summarize, the analysis presented in this paper contains three main problems that restrict the scope of the conclusions, i.e. (1) no experimental assignment of the selection methods to the country-surveys, (2) mixing of the impact of sampling procedures with the effects of other factors, and (3) the character of birthday sampling as not purely random.

4.2. Methods for assessing the impact of within-household selection procedures on the values of refusal and cooperation rates

In order to verify the first hypothesis (H1), refusal rates and cooperation rates were calculated for each of the 98 surveys included in the meta-analysis. More precisely, each individual sample / study was characterized by the first version of the refusal

rate (REF1) and the second version of the cooperation rate (COOP2), where the definitions of REF1 and COOP2 are specified by the American Association for Public Opinion Research (AAPOR, 2016). REF1 is the proportion of all eligible cases in which a household member or the target person refuses to be interviewed or breaks off an interview, while COOP2 is the number of all partial and completed interviews divided by the number of all the units with whom contact was made.

Analysis of ESS data has been conducted to assess the following comparisons of within-household selection procedures: (1) Kish grid vs. any of the birthday methods (last / next / closest), (2) Kish grid vs. last-birthday method, (3) Kish grid vs. next-birthday method, (4) last-birthday method vs. next-birthday method. It should be mentioned that only 5 studies utilized the closest-birthday procedure. Thus, this method was excluded from comparisons between different birthday methods, however, it was included in the comparisons between Kish grid samples and the studies where any of the birthday methods was utilized. The size of the impact of within-household selection methods on REF1 and COOP2 were calculated as the odds ratios, i.e., for REF1:

$$OR_{REF1} = \frac{\overline{REF1}_{method1} \cdot (1 - \overline{REF1}_{method2})}{\overline{REF1}_{method2} \cdot (1 - \overline{REF1}_{method1})},$$

and for COOP2:

$$OR_{COOP2} = \frac{\overline{COOP2}_{method1} \cdot (1 - \overline{COOP2}_{method2})}{\overline{COOP2}_{method2} \cdot (1 - \overline{COOP2}_{method1})},$$

where $\overline{REF1}$ and $\overline{COOP2}$ are respectively the weighted means of single-study values of REF1 and COOP2 obtained under a specific within-household selection procedure; weights are computed as reciprocals of single-study variance of REF1 and COOP2 estimators. In other words, odds ratios rely on a comparison of the weighted mean odds of obtaining refusal (or cooperation) under different procedures of within-household selection. An odds ratio of 1.0 means that there is no effect at all, i.e., the odds of obtaining refusal (or cooperation) are equal under both these procedures. The larger the deviation from 1.0 the larger the effect. An odds ratio above 1.0 indicates that obtaining the refusal or cooperation is more likely to occur if the first method of within-household selection is implemented, while an odds ratio below 1.0 indicates that the event is less likely to occur under the first procedure of within-household selection.

4.3. Assessing demographic representation of survey samples with external benchmark data

The impact of Kish grid and birthday procedures on gender and age imbalance (hypothesis H2) has been analysed according to the idea of comparing survey results with more accurate sources of external benchmark data, i.e. the so-called gold standards. As concluded by Groves (2006), the strengths of this method are that one can compare estimates independent of survey questions and no information is required at an individual level. By contrast, the main weakness is that key survey estimates do not usually exist in external sources. However, when a source of external data enjoys great credibility among users, as would for example a high-quality government population database, then obtaining results similar to those in the external data increases confidence about survey quality.

A comparison between survey estimates and external benchmark data is possible if variables have been measured in an identical way or at least if measurements could be recoded to a common standard. Koch (2016), and earlier Koch et al. (2014), examined a demographic representation of ESS studies (rounds 5 and 6) by comparing the survey distributions of gender, age, marital status, work status, nationality and household size, to the distributions of same variables obtained in the European Union Labour Force Survey (LFS). It may be noted, however, that LFS is a survey with serious non-response and measurement problems in some countries (Eurostat 2014). In this paper only Eurostat census register data on gender and age were considered as gold standards for ESS results (Table 2).

Table 2. Variables of the ESS and Eurostat comparison

Variable	ESS variable	Eurostat variable ^a	ESS survey estimator ^b	Eurostat population parameter
Gender	gndr	demo pjangroup sex	% of females	% of females in population 15+
Age: 15–24 years old	agea (recoded)	demo pjangroup age	% of respondents aged 15–24 years old	% of population aged 15–24 in population 15+
Age: 75 years and older	agea (recoded)	demo pjangroup age	% of respondents 75 years and older	% of population aged 75+ in population 15+

^a Eurostat *demo pjangroup* data was not available for the following ESS studies: Albania (ESS6); Israel (ESS1, ESS4, ESS5, ESS6, ESS7) and Kosovo (ESS6). In such cases information about population structure was drawn from appendices to the ESS Documentation Reports (ESS Round 1, 2016; ESS Round 4, 2016; ESS Round 5, 2016; ESS Round 6, 2016; ESS Round 7, 2016).

^b ESS data were weighted by design weights (i.e., *dweight* variable in cumulative dataset).

It should be borne in mind that in the ESS project only informants of 15-years and over were included in the target population. Thus, in order to compare ESS results with the external population data the informants aged 0-14 had to be removed from the census database. It is worth noting that the Eurostat *demo pjangroup* variable (i.e., the gold standard for ESS results) provides the values for population size in five-year age groups and allows extraction of the cohort of people aged 0-14. However, the main problem with Eurostat data lies in the timing of the fieldwork execution of ESS studies, since the *demo pjangroup* variable involves counts of the population for the 1st of January, while in the ESS project the fieldwork has usually extended over a period of several months. Thus, in order to render Eurostat data comparable to those observed in the ESS project the decision was as follows: (1) if ESS fieldwork took place in the middle of a year, the reference date for Eurostat data was the 1st of January of that year, (2) if fieldwork took place at the turn of the year, the reference date was the 1st of January of the year when ESS fieldwork was finished.

Three measures of differences between ESS results (i.e., values of estimators denote as \hat{p}_i) and Eurostat benchmark data (i.e., true values of parameters denote as p_i) were calculated for each survey included in the analysis to verify hypothesis H2:

(1) the measure of under- or overrepresentation of women in the survey:

$$\% \text{ females in study} - \% \text{ of females in population aged 15+}$$

(2) the measure of under- or overrepresentation of young people in the survey:

$$\% \text{ of respondents aged 15–24 in study} - \% \text{ of population aged 15–24} \\ \text{in population aged 15+}$$

(3) the measure of under- or overrepresentation of old people in the survey:

$$\% \text{ of respondents aged 75+ in study} - \% \text{ of population aged 75+} \\ \text{in population aged 15+}.$$

The differences between estimator and parameter values that were calculated for each study are referred to as ‘single-study effect size’ in the later meta-analysis. As the true proportions of gender and age groups are known from the Eurostat database, the within-study variance of each estimator is equal to $V_i = p_i(1 - p_i)/n_i$, where p_i is a true fraction of gender and age categories in the population of informants of 15 years of age and older, and n_i is the total number of respondents in each of the 98 different surveys included in meta-analysis.

4.4. Assessing the degree of interviewers' influence on within-selection processes by means of internal criteria of representativeness

The impact of Kish grid and birthday procedures on the degree of interviewers' impact on the within-household selection process (hypothesis H3) has been analysed according to the method proposed originally by Sodeur (1997). Kohler (2007) developed Sodeur's idea of the so-called internal criteria of representativeness and applied this method to a comparison of the quality of the sampling processes in different international survey programmes. The basic idea of internal criteria is to measure the size of the sampling bias for a subgroup of the entire sample for which the true parameter value is known, e.g. for a subgroup of respondents living in households with heterosexual couples where only one of the partners is interviewed. In such households, both male and female respondents have the same probability of being selected. Thus, if there are no significant deviations in the sampling, none of the genders is expected to be significantly under- or overrepresented in the survey. One of the main advantages of internal criteria of representativeness lies in the possibility of comparing different surveys without the necessity of collecting any external data. Moreover, the 50/50 gender ratio of heterosexual couples is not affected by any sort of measurement errors and is unaffected by household size (Kohler, 2007).

Nevertheless, deviations from the true gender ratio may be affected by differences in refusal behaviour, since females are more reluctant to cooperate than males (Stoop, 2004, 2005). Menold (2014) used the observation about gender differences in refusals to demonstrate that the representation of females exceeding 50% in a subsample of heterosexual couples may plausibly be explained by interviewers' influence on the selection process (i.e., illegal substitution of another unit for a nonresponding one), while female representation below 50% may plausibly be explained by greater reluctance to participate in surveys. In other words, both over- and underrepresentation of females in a subgroup of respondents as representatives of heterosexual couples indicates distortion of the selection process, however, only an overrepresentation of women indicates the presence of interviewers' influence, while an underrepresentation of women is a consequence of lower cooperation rates among female respondents.

The implementation of internal criteria of representativeness required extraction of a subsample of heterosexual couples from the entire cumulative ESS1-ESS7 dataset. Such a subsample contains respondents living together within one household with a partner of the opposite sex, on condition that both partners belong to the target population. As was mentioned earlier, the ESS target population contains all persons who are 15 years of age or over, residing in private households. Therefore, singles, or partners not living together within one household, homosexual partners

and partners living together with other relatives and/or non-relatives who belong to the target population, e.g. parents, parents-in-law and children of 15 years of age and over, were excluded from the entire ESS1-ESS7 sample. The subsample separation process is described in Table 3.

Table 3. Basic characteristics of ESS1-ESS7 subsample separation process

Reasons for excluding respondents from ESS1-ESS7 cumulative dataset	number of units
Step 1: One-person households or households with unknown number of members	42,305
Step 2: Respondent does not live with husband/wife/partner or there is no information about it	43,317
Step 3: Respondent lives with: a) parents/parents-in-law, b) other relative, c) other non-relative	9,575
Step 4: Respondent lives with a homosexual partner	1,018
Step 5: Respondent declared cohabitation in one household with two or more husbands/wives	119
Step 6: No data on: a) gender of respondent/partner, b) relationship with other residents	820
Step 7: Respondent lives in a household with children aged 15 years and over	26,428
Total number of respondents excluded from ESS1-ESS7 cumulative dataset	123,582

A subsample of $n = 73,989$ respondents living with a partner of the opposite sex (with or without children aged 0–14) has been selected from the entire ESS1-ESS7 dataset. This means that the extraction process has reduced the total number of 197,565 units in a cumulative database to less than 50%. However, even if the extraction process has reduced the size of the sample, it was the only way to calculate for each survey included in our analysis the difference (called in meta-analysis the ‘single-study effect size’) between the observed proportion of females in households of heterosexual partners (denote as \hat{p}_i) and the expected proportion of females in such households, i.e.

$$\% \text{ of females in a subsample of study} - 50\%.$$

As the expected proportion of females is equal to p_i , the within-study variance of the female ratio estimator is equal to $V_i = 0.25/n_i$, where n_i is the total number of respondents in each subsample extracted from 98 different surveys included in the meta-analysis of ESS studies (Kohler 2007).

4.5. Overall effect size as a quantitative measure of the impact of the Kish grid and birthday methods by means of meta-analysis

Baker (2007) distinguished three main meanings of the term ‘meta-analysis’ in social science, i.e. (1) a quantitative literature review of a body of empirical findings, (2) a summary of replication research on a specific topic and (3) a theoretical analysis of the complex philosophical problems associated with commonalities in scientific approaches. The first and the second meaning of the term ‘meta-analysis’ are associated with a specific type of research design. One can also note that the main purpose of meta-analysis is to involve the results of different studies, as a unit of analysis, and to combine single-study estimators (the so-called ‘effect sizes’) into one measure called the overall (average weighted) effect size.

In the last three decades, meta-analysis has been explored extensively mainly in medicine and psychology. However, many examples of meta-analysis research design can also be found in survey research methodology. In the latter studies, meta-analysis was mainly used to evaluate mode-effects or sampling-effects (Hox & de Lee, 1994; Manfreda, Bosnjak, Berzelak, Hass, & Vehovar, 2008; Medway & Fulton, 2012; Shih & Fan, 2007; Yan, 2009), relations between response rate and non-response bias (Groves, 2006; Groves & Peytcheva, 2008), effectiveness of advance letters and incentives in increasing response rates (Church, 1993; Hopkins & Gullickson, 1992; de Leeuw, Callegaro, Hox, Korendijk, & Lensvelt-Mulders, 2007) as well validity and reliability of survey questions (Scherpenzeel & Saris, 1997).

In this paper, meta-analysis has not been used as a type of research design but only as an analytical procedure that allows achievement of a correct estimation of overall effect size on the basis of different surveys from the same research project. The main goal was to quantitatively summarize the results of primary ESS surveys by considering both within- and between-study variance of single study effect sizes. Nevertheless, even if the use of meta-analysis was quite specific, one can note that the five major stages that are typical for meta-analysis research design have been implemented in the analysis of the ESS. These are the following steps: (1) formulation of the research problem, (2) collection of primary surveys, (3) evaluation of the quality of primary surveys, (4) analysis of the meta-analytic data, and (5) presentation of the meta-analytic results (Card, 2012).

To compute a quantitative measure of the impact of Kish grid and birthday methods on demographic representation and of interviewers’ influence on the selection process a concept of the overall effect size in a form proposed by Borenstein et al. (2009) was used. As was described in the above sections of this article the single study effect size measures were calculated separately for each ESS survey. It should be emphasized that the so-called single-study effect size

amounts to the difference between survey estimate \hat{p}_i and true parameter value p_i , i.e.,

$$ES_i = \hat{p}_i - p_i.$$

Thus, the definition of ES_i is consistent with the well-known concept of total survey error (Biemer, 2010). As was mentioned, the main goal of meta-analysis is to integrate the values of single-study effect sizes into one measure called the ‘overall effect size’.

One may note that if all the studies included in meta-analysis had been equally precise, the overall effect size could have been calculated as a simple average (mean) of single-study effect sizes. However, in survey practice this assumption is rarely satisfied; one should calculate the overall effect size rather than a weighted mean. Borenstein et al. (2009) proposed two conceptually different approaches to estimating values of the overall (weighted) effect size. One of these approaches is the so-called fixed effect model, while the second is the random effect model. The fixed effect model assumes one true effect size for all studies included in an analysis. This assumption means that all factors which could influence the effect size are the same in all studies, and single-study effect sizes differ only because of the random error within each survey. Thus, to calculate the overall effect size one should incorporate weights as reciprocals of within-study variance. By contrast, the random effect model assumes that no one true effect size exists, but instead multiple true effects may be entailed in each study. The latter model is appropriate when different factors influence true effects in each survey. In consequence, measures of single-study effect sizes do not calculate one true effect size, but estimate different true effects instead. Values of single-study effect sizes vary from study to study not only because of within-study variance, but also as a result of between-study variance. Thus, the sum of within- and between-study variances should be incorporated in weights.

The random effect model of meta-analysis is much more conservative than the first model, i.e., much higher deviations from underlying true parameters must be attested if they are to be recognized as statistically significant. This observation serves as the basis for recognizing the overall effect size as statistically significant. However, the random effect model is much more accurate in meta-analysis of ESS studies, since each survey provides information about effects in different populations and each survey differs in the sampling process and fieldwork procedures. Thus, in a random effect model the overall effect size is computed as a weighted mean, i.e.,

$$\overline{ES}^* = \frac{\sum_{i=1}^k w_i^* \cdot ES_i}{\sum_{i=1}^k w_i^*}$$

where k is the number of studies included in meta-analysis and w_i^* denotes weights assigned to each study.

Weights w_i^* are computed as reciprocals of $V_i^* = V_i + \tau^2$, where V_i^* is the sum of within-study variance (denoted as V_i) and between-study variance (denoted as τ^2). Borenstein et al. (2009) defined τ^2 as a maximum value from a set of two values $\left\{\frac{Q-df}{c}; 0\right\}$, where:

- $Q = \sum_{i=1}^k (ES_i - \overline{ES})^2 / V_i$, whereby $\overline{ES} = \sum_{i=1}^k V_i^{-1} \cdot ES_i / \sum_{i=1}^k V_i^{-1}$;
- $df = k - 1$;
- $c = \sum_{i=1}^k V_i^{-1} - \sum_{i=1}^k V_i^{-2} / \sum_{i=1}^k V_i^{-1}$.

The variance of overall effect size \overline{ES}^* is then computed as $V^* = 1 / \sum_{i=1}^k w_i^*$, while the standard error of \overline{ES}^* is computed as a square root of variance V^* , i.e., $SE(\overline{ES}^*) = \sqrt{V^*}$. Based on the standard error of \overline{ES}^* one can also compute the confidence interval for \overline{ES}^* , as well as the z -value and p -value to verify whether the overall effect size is statistically significant at a fixed level of significance.

5. RESULTS

5.1. Comparison of the impact of Kish grid and birthday procedures on survey outcome rates

The odds ratios for obtaining refusals or cooperation were analysed to test hypothesis H1 which predicts lower refusal rates and higher cooperation rates in studies where birthdays were used in the sampling process. One can note that in a group of ESS studies where the Kish grid was utilized the mean value of the refusal rate (mean REF1 equal to 25.5%) was higher by 5pp. than the mean value of the refusal rate in a group of ESS studies where birthday methods were implemented (mean REF1 equal to 20.1%). Furthermore, in a group of Kish grid samples the mean value of COOP2 (equal to 66.4%) was lower by 4pp. than the mean value of the cooperation rate in a group of birthday samples (mean COOP2 equal to 70.1%). It should also be noted that there were no differences between mean values of cooperation rates obtained in last- and next-birthday samples, i.e., in both cases the mean value of COOP2 was equal to 70.1%. However, last- and next-birthday procedures differ in values of refusal rates, since the mean value of REF1 in a group of last-birthdays studies was equal to 19.9%, while in a group of next-birthday studies the mean REF1 was equal to 22.8%.

In order to verify whether these differences are significant, the odds ratios were calculated (Table 4).

Table 4. Impact of Kish grid and birthday procedures on refusal rates and cooperation rates

	odds ratio	
	REF1	COOP2
Kish grid vs. any birthday method	1.362*	0.824*
Kish grid vs. Last-birthday method	1.369*	0.837*
Kish grid vs. Next-birthday method	1.159*	0.839*
Last-birthday vs. Next-birthday method	0.847*	1.002

* p -value < 0.001

The results of the analysis confirm that in the Kish grid samples the odds of obtaining cooperation were significantly lower while the odds of obtaining refusals were significantly higher. In addition, there were no significant differences in the odds of obtaining cooperation in last- and next-birthday procedures, however the odds of obtaining refusals were significantly lower in last-birthday samples.

Concerning hypothesis H1, it may be observed that the implementation of birthday methods, compared to the implementation of the Kish grid, results in higher cooperation rates and lower refusal rates.

5.2. Comparison of the impacts of Kish grid and birthday procedures on demographic representation

The impact of Kish grid and birthday methods on gender and age imbalance has been analysed to verify hypothesis H2. This hypothesis predicts that both Kish grid and birthday methods have a significant impact on gender and age imbalance. Hypothesis H2 also predicts a significant overrepresentation of females and a significant underrepresentation of younger and older respondents both in Kish grid and birthday samples. However, the size of this impact is expected to be lower in Kish grid samples.

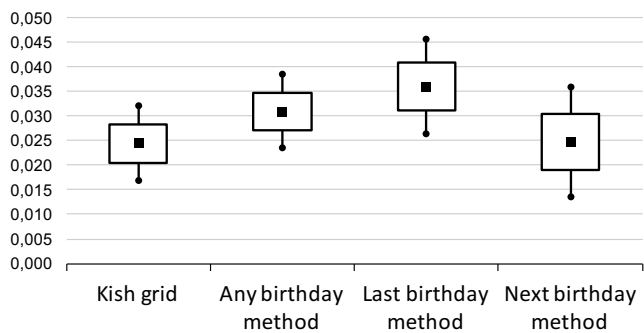
Table 5 presents measures of the overall impact of within-household selection procedures on female representation. The mean values of overall effect size, standard errors and z -values were calculated for each procedure of within-household selection. Additionally, Figure 1 presents the mean values of overall effect size (points) as well as the respective values of standard errors (boxes) and 95% confidence intervals of such a mean (whiskers) for each in order to compare the procedures of within-household selection.

Table 5. Impact of Kish grid and birthday procedures on proportion of females

Within-household procedure	number of studies	Overall effect size		
		\overline{ES}^*	$SE(\overline{ES}^*)$	z-value
Probability method				
Kish grid	34	0.024	0.004	6.33***
Quasi-probability methods				
Any of birthday method	61	0.031	0.004	8.09***
Last-birthday method	40	0.036	0.005	7.26***
Next-birthday method	17	0.025	0.006	4.33***

*** p -value < 0.001

Figure 1. Box-and-whisker-plots of overall effect size (proportion of females) by type of within-household selection procedure



Several trends can be noticed based on Table 5 and Figure 1. Firstly, it may be observed that each method of within-household selection has a significant impact on the imbalance of female participation. More precisely, Kish grid samples tend to overrepresent women by 2.4 pp., while the class of birthday methods tends to overrepresent women by 3.1 pp. Secondly, one may notice significant differences within the class of birthday methods, since last-birthday samples tend to overrepresent women much more than next-birthday samples. Finally, the standard errors of the overall mean effect size are lower for Kish grid studies than for the last- and next-birthday samples.

The criteria of analysis that were applied to assess the impact of within-household selection procedures on gender representation were also used to investigate the

impact of Kish grid and birthday methods on age imbalance. Table 6 and Figure 2 present the impact of within-household selection methods on the proportion of respondents aged 15–24, while Table 7 and Figure 3 present the impact of such procedures on the proportion of respondents aged 75 years and older.

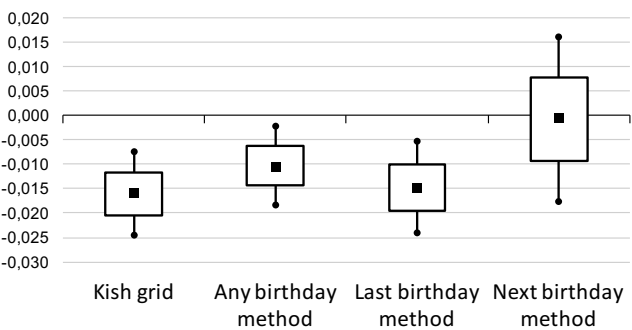
Table 6. Impact of Kish grid and birthday procedures on the proportion of people aged 15–24

Within-household procedure	number of studies ^a	Overall effect size		
		\overline{ES}^*	$SE(\overline{ES}^*)$	z-value
Probability method				
Kish grid	34	-0.016	0.004	-3.66 [*]
Quasi-probability methods				
Any of birthday method	59	-0.010	0.004	-2.51 ^{**}
Last-birthday method	40	-0.015	0.005	-3.09 ^{**}
Next-birthday method	15	-0.001	0.009	-0.09

^{*} p -value < 0.001
^{**} p -value < 0.05

^a Data from Austria (ESS2, ESS3; last-birthday samples) were excluded from analysis, since the overrepresentation of respondents aged 15-24 was twice the number found in other studies. These values were extreme outliers which might have significant influence on meta-analysis results.

Figure 2. Box-and-whisker-plots of overall effect size (proportion of people aged 15–24) by type of within-household selection procedure



It may be noted that both the Kish grid and birthday methods tend to underrepresent younger respondents. However, when one splits the class of birthday procedures into different methods, the impact of the last-birthday method

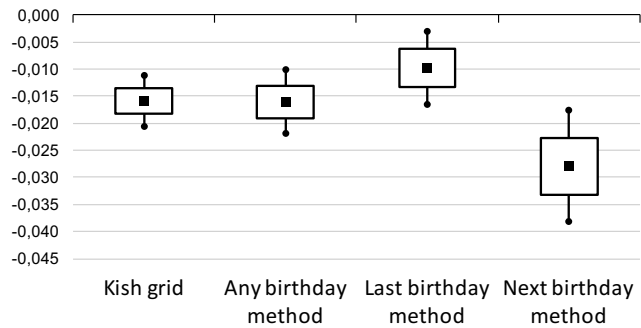
on the underrepresentation of young respondents is significant, while the impact of the next-birthday method is negligible.

Table 7. Impact of Kish grid and birthday procedures on proportion of people aged 75+

Within-household procedure	number of studies	Overall effect size		
		\overline{ES}^*	$SE(\overline{ES}^*)$	z-value
Probability method				
Kish grid	34	-0.016	0.002	-6.65 [*]
Quasi-probability methods				
Any of birthday method	61	-0.016	0.003	-5.38 [*]
Last-birthday method	40	-0.010	0.003	-2.84 ^{**}
Next-birthday method	17	-0.028	0.005	-5.31 [*]

^{*} p-value < 0.001
^{**} p-value<0.05

Figure 3. Box-and-whisker-plots of overall effect size (proportion of people aged 75+) by type of within-household selection procedure



It transpires from Table 7 and Figure 3 that each of the procedures of within-household selection that were subject to comparison have significant impact on the overrepresentation of respondents aged 75+. Minor differences in the mean overall effect size were observed between the Kish grid and the class of birthday methods; however, next-birthday samples tend to underrepresent older respondents when compared with the last-birthday samples, and by more than twice as much.

Summarizing the results presented above it may be stated that hypothesis H2 is only partially corroborated. As predicted by this hypothesis, the implementation of

Kish grid and birthday procedures results in an overrepresentation of females and underrepresentation of younger and older respondents. However, contrary to the predictions described in hypothesis H2, the scale of this impact is not significantly lower in Kish grid samples.

5.3. Comparison of the impact of Kish grid and birthday procedures on the degree of interviewers’ influence on the within-household selection process

The final section presents the results of meta-analysis for testing hypothesis H3. This hypothesis predicts that the impact of within-household selection procedures on the degree of interviewers’ influence on the within-household selection differs according to which selection method has been applied. If birthday methods have more problems with accuracy in selecting target respondents, then deviations from the true 50/50 gender ratio in a subsample of respondents as representatives of heterosexual couples should be higher in birthday samples.

Table 8 displays the mean values of overall effect size, the standard errors of such means and z-values for each type of within-household selection procedure. In turn, Figure 4 displays box-and-whisker-plots with mean overall effect size (points), standard error of mean (boxes) and 95% confidence intervals of such a mean (whiskers).

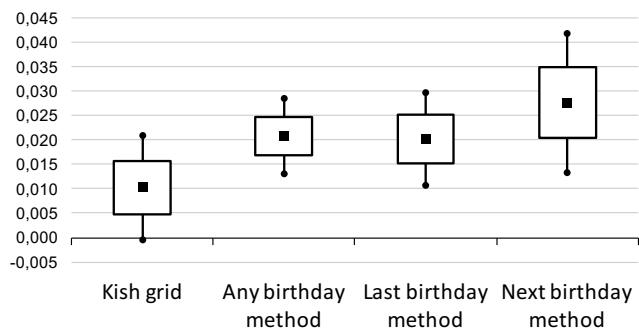
Table 8. Impact of Kish grid and birthday procedures on proportion of females in households of heterosexual partners

Within-household procedure	number of studies ^a	Overall effect size		
		\overline{ES}^*	$SE(\overline{ES}^*)$	z-value
Probability method				
Kish grid	31	0.010	0.006	1.88
Quasi-probability methods				
Any of birthday method	60	0.021	0.004	5.29 [*]
Last-birthday method	39	0.020	0.005	4.12 [*]
Next-birthday method	17	0.028	0.007	3.80 [*]

^{*} p-value < 0.001

^a Data from Slovakia (ESS4, ESS5, ESS6; Kish grid samples) and Ukraine (ESS5; last-birthday sample) were excluded from analysis, since the values for overrepresentation of women were extreme outliers which might have significant influence on meta-analysis results.

Figure 4. Box-and-whisker-plots of overall effect size (proportion of females in households of heterosexual partners) by type of within-household selection procedure



The results of meta-analysis support the expectations of hypothesis H3. This means that deviations from the true 50/50 gender ratio parameter in households of heterosexual partners vary depending on the type of within-household selection procedure employed. As was expected, birthday methods (examined as one class of birthday methods and divided into last- and next-birthday method) have a significant impact on interviewers’ influence on the within-household selection, while this impact is reduced twofold (and is even negligible) in Kish grid samples. In other words, both the Kish grid and birthday samples tend to overselect women, however the implementation of the Kish grid reduces the chances of selecting women in households of heterosexual partners and improves selection quality. It should be emphasized that the overrepresentation of women indicates the presence of interviewers’ impact on the selection process, which is caused by overselecting of the informants who stay at home more than others.

6. CONCLUSIONS

The findings presented in this paper are instrumental in developing an understanding of the risks of utilizing different types of within-household selection procedures in face-to-face surveys based on probability samples. Firstly, as expected, in the Kish grid method the full enumeration of all eligible individuals comprising the household has a negative impact on cooperation and increases the number of refusals to participate in the survey. Thus, in Kish grid samples, the refusal rates are significantly higher, while the cooperation rates are significantly lower compared to those in birthday samples. Secondly, it was demonstrated that both Kish grid and birthday methods have a significant impact on gender and age imbalance. Meta-analysis confirmed that Kish grid and birthday methods result

in an overrepresentation of women and underrepresentation of younger and older respondents. Finally, it was demonstrated that birthday methods involve more problems with accuracy of the within-household selection process. The latter means that the risk of significant interviewers' influence on the selection process constitutes a serious problem especially in birthday samples, where interviewers have more opportunity to substitute another unit from the same household for a nonresponding unit.

In conclusion, the analyses presented in this paper indicate that both Kish grid and birthday procedures have a significant impact on fieldwork and sample quality. The main advantages of birthday procedures lie in the ease of implementation of the sampling process, which increases cooperation and decreases refusals. However, the cost of the birthday procedures is usually an incomplete randomness in the selection. On the other hand, the main advantage of the Kish grid lies in possibility of producing consistent and unbiased estimates, but at the same time the cost is a decrease in cooperation and an increase in refusals. Nevertheless, one should bear in mind that the decision as to which type of within-household selection procedure to use in a sampling process constitutes a major challenge, therefore one should consider the total survey error produced by each of the within-household selection methods.

REFERENCES

- Bakker, J., I., (Hans). (2007). Meta-analysis. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology* (pp. 2963–2964). Malden: Blackwell Publishing Ltd.
- Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. Hoboken: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470891056>
- Biemer, P. P. (2010). Total Survey Error. Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74(5), 817–848. <https://doi.org/10.1093/poq/nfq058>
- Binson, D., Canchola, J. A., & Catania, J. A. (2000). Random Selection in a National Telephone Survey: A Comparison of the Kish, Next-Birthday, and Last-Birthday Methods. *Journal of Official Statistics*, 16(1), 53–59.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Hoboken: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470743386>
- Brogan, D. J., Denniston, M. M., Liff, J. M., Flagg, E. W., Coates, R. J., & Brinton, L. A. (2001). Comparison of telephone sampling and area sampling: Response rates and within-household coverage. *American Journal of Epidemiology*, 153(11), 1119–1127. <https://doi.org/10.1093/aje/153.11.1119>
- Brooks, C. A., & Bailar, B. (1978). *Statistical Policy Working Paper 3 – An Error Profile: Employment as Measured by the Current Population Survey*. Washington: Office of Federal Statistical Policy and Standards, U.S. Department of Commerce.
- Bryant, B. E. (1975). Respondent Selection in a Time of Changing Household Composition. *Journal of Marketing Research*, 12(2), 129–135. <https://doi.org/10.2307/3150434>

- Card, N. A. (2015). *Applied meta-analysis for social science research*. New York: Guilford Publications.
- Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, 57(1), 62–79. <https://doi.org/10.1086/269355>
- Czaja, R., Blair, J., & Sebestik, J. P. (1982). Respondent Selection in a Telephone Survey: A Comparison of Three Techniques. *Journal of Marketing Research*, 19(3) 381–385. <https://doi.org/10.2307/3151573>
- Denk, C. E., & Hall, J. W. (2000). *Respondent Selection in RDD Surveys: A Randomized Trial of Selection Performance*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Portland, OR.
- Denk, C. E., Guterbock, T. M., & Gold, D. B. (1996). *Modelling Selection of Respondents Within Household in Telephone Surveys*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Salt Lake City, UT.
- ESS Round 1: European Social Survey (2016): *ESS-1 2002 Documentation Report. Edition 6.5*. Bergen, European Social Survey Data Archive, NSD – Norwegian Centre for Research Data for ESS ERIC.
- ESS Round 4: European Social Survey (2016): *ESS-4 2008 Documentation Report. Edition 5.4*. Bergen, European Social Survey Data Archive, NSD – Norwegian Centre for Research Data for ESS ERIC.
- ESS Round 5: European Social Survey (2016): *ESS-5 2010 Documentation Report. Edition 4.1*. Bergen, European Social Survey Data Archive, NSD – Norwegian Centre for Research Data for ESS ERIC.
- ESS Round 6: European Social Survey (2016): *ESS-6 2012 Documentation Report. Edition 2.3*. Bergen, European Social Survey Data Archive, NSD – Norwegian Centre for Research Data for ESS ERIC.
- ESS Round 7: European Social Survey (2016): *ESS7- 2014 Documentation Report. Edition 3.1*. Bergen, European Social Survey Data Archive, NSD – Norwegian Centre for Research Data for ESS ERIC.
- Eurostat (2014). *Quality report of the European Union Labour Force Survey 2013*. Eurostat Statistical Working Papers.
- Gaziano, C. (2005). Comparative Analysis of Within-Household Respondent Selection Techniques. *Public Opinion Quarterly*, 69(1), 124–157. <https://doi.org/10.1093/poq/nfi006>
- Groves, R. M. (1989). *Survey Errors and Survey Costs*, Hoboken: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471725277>
- Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646–675. <https://doi.org/10.1093/poq/nfl033>
- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72(2), 167–189. <https://doi.org/10.1093/poq/nfn011>
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. Hoboken: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118490082>
- Groves, R. M., & Kahn, R., L. (1979). *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.

- Hagan, D. E., & Collier C., M. (1983). Must Respondent Selection Procedures for Telephone Surveys Be Invasive? *Public Opinion Quarterly*, 47(4), 547–556. <https://doi.org/10.1086/268811>
- de Heer, W. (1999). International Response Trends: Results of an International Survey. *Journal of Official Statistics*, 15(2), 129–142.
- Hopkins, K. D., & Gullickson, A. R. (1992). Response rates in survey research: A meta-analysis of the effects of monetary gratuities. *Journal of Experimental Education*, 61(1), 52–62. <https://doi.org/10.1080/00220973.1992.9943849>
- Hox, J. J., & de Leeuw, E. (1994). A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys. Applying Multilevel Modeling to Meta-Analysis. *Quality and Quantity*, 28(4), 329–344. <https://doi.org/10.1007/BF01097014>
- Keeter, S., & Fisher, K. (1997). *A Comparison of 'Last Birthday' and 'Youngest Male/Oldest Female' Respondent Selection Procedures*. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Petersburg, FL.
- Kish, L. (1949). A Procedure for Objective Respondent Selection within the Household. *Journal of the American Statistical Association*, 44(247), 380–387. <https://doi.org/10.1080/01621459.1949.10483314>
- Koch, A. (2016). *Assessment of socio-demographic sample composition in ESS Round 6*. Mannheim: European Social Survey, GESIS.
- Kohler, U. (2007). Survey from Inside: An Assessment of Unit Nonresponse Bias With Internal Criteria. *Survey Research Methods*, 2(1), 55–67.
- Korns, A. (1977). Coverage Issues Raised by Comparisons Between CPS and Establishment Employment. In American Statistical Association, *Proceedings of the Social Statistics Section, Part I* (pp. 60–69).
- Lavrakas, P. L., Stasny, E. A., & Harpuder, B. (2000). A Further Investigation of the Last-Birthday Respondent Selection Method and Within-Unit Coverage Error. In American Statistical Association, *Proceedings of the Survey Research Methods Section* (pp. 890–895).
- Lavrakas, P. L., Bauman, S. L., & Merkle, D. M. (1993). The Last-Birthday Selection Method & Within-Unit Coverage Problems. In American Statistical Association, *Proceedings of the Survey Research Methods Section* (pp. 1107–1112).
- Le, K. T., Brick, M., Diop, A., & Alemadi, D. (2013). Within-household Sampling Conditioning on Household Size. *International Journal of Public Opinion Research*, 25(1), 108–118. <https://doi.org/10.1093/ijpor/eds008>
- de Leeuw, E. D., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys: A meta-analysis. *Public Opinion Quarterly*, 71(3), 413–443. <https://doi.org/10.1093/poq/nfm014>
- Lynn, P., Gabler, S., Häder, S., & Laaksonen, S. (2007). Methods for Achieving Equivalence of Samples in Cross-National Surveys: The European Social Survey Experience. *Journal of Official Statistics*, 27(1), 107–124.
- Manfreda, K. L., Bosnjak, M., Berzelak, J., Hass, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 50(1), 79–104.
- Martin, E. (1999). Who Knows Who Lives Here? Within-Household Disagreements as a Source of Survey Coverage Error. *Public Opinion Quarterly*, 63(2), 220–236. <https://doi.org/10.1086/297712>

- Medway, R. L., & Fulton, J. (2012). When More Gets You Less: A Meta-analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates. *Public Opinion Quarterly*, 76(4), 733–746. <https://doi.org/10.1093/poq/nfs047>
- Menold, N. (2014). The Influence of Sampling Method and Interviewers on Sample Realization in the European Social Survey. *Survey Methodology*, 40(1), 105–123.
- O'Rourke, D., & Lakner, E. (1989). Gender bias: Analysis of factors causing male underrepresentation in surveys. *International Journal of Public Opinion Research*, 1(2), 164–176. <https://doi.org/10.1093/ijpor/1.2.164>
- O'Rourke, D., and Blair, J. (1983). Improving Random Respondent Selection in Telephone Surveys. *Journal of Marketing Research*, 20(4), 428–432. <https://doi.org/10.2307/3151446>
- Oldendick, R. W., Bishop, G. F., Sorenson, S. B. & Tuchfarber, A. F. (1988). A Comparison of the Kish and Last Birthday Methods of Respondent Selection in Telephone Surveys. *Journal of Official Statistics*, 4(4), 307–318.
- Paisley, W. J., & Parker, E. B. (1965). A Computer-Generated Sampling Table for Selecting Respondents Within Households. *Public Opinion Quarterly*, 29(3), 431–436. <https://doi.org/10.1086/267343>
- Salmon, C. T., & Nichols, J. S. (1983). The Next-Birthday Method of Respondent Selection. *Public Opinion Quarterly*, 47(2), 270–276. <https://doi.org/10.1086/268785>
- Shih, T., & Fan, X. (2007). Response Rates and Mode Preferences in Web-mail Mixed Mode Surveys: A Meta-analysis. *International Journal of Internet Science*, 2(1), 59–82.
- Smith, W. (2007). Survey Non-Response Procedures in Cross-National Perspective: The 2005 ISSP Non-Response Survey. *Survey Research Methods*, 1(1), 45–54.
- Smith, W., Chey, T., Jalaludin, B., Salkeld, G., & Capon, T. (1995). Increasing Response Rates in Telephone Surveys: A Randomized Trial. *Journal of Public Health*, 17(1), 33–38.
- Sodeur, W. (1997). Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 41, 58–82.
- Stoop, I., Billiet, J., Koch, A., & Fitzgerald R. (2010). *Improving Survey Response. Lessons learned from the European Social Survey*. Hoboken: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470688335>
- Stoop, I. (2004). Surveying Nonrespondents. *Field Methods*, 16(1), 23–54. <https://doi.org/10.1177/1525822X03259479>
- Stoop, I. (2005). *The Hunt for the Last Respondent. Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office.
- Scherpenzeel, A., & Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods and Research*, 25 (3), 341–383. <https://doi.org/10.1177/0049124197025003004>
- The American Association for Public Opinion Research (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition*. AAPOR.
- Tourangeau, R., Shapiro, G., Kearney, A., & Ernst, L. (1997). Who lives here? Survey undercoverage and household roster questions. *Journal of Official Statistics*, 13(1), 1–18.

- Troldahl, V. C., & Carter Jr, R. E. (1964). Random Selection of Respondents Within Household in Phone Surveys. *Journal of Marketing Research*, 1(2), 71–76. <https://doi.org/10.2307/3149926>
- Valentine, C. A., & Valentine, B. L. (1971). *Missing Men: A Comparative Methodological Study of Underenumeration and Related Problems*. Washington: The U.S. Census Bureau.
- Yan, T. (2009). *A Meta-analysis of Within-Household Respondent Selection Methods*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Hollywood, FL.
- Ziniel, S. (2008). Within-Unit Coverage Error. In P. Lavrakas (Ed.), *Encyclopaedia of Survey Research Methods* (pp. 962–963). Thousand Oaks: SAGE Publications, Inc.

Piotr Jabkowski is an Associate Professor at the Institute of Sociology of Adam Mickiewicz University in Poznan. His research interests include problems related to the quality of survey data, including the total survey error paradigm and sampling theory.

E-mail: pjabko@amu.edu.pl